



Desafíos metodológicos e innovaciones en el estudio de las élites

Bastián González-Bustamante

University of Oxford

Universidad de Santiago de Chile

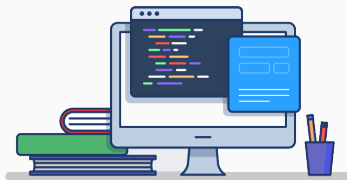
✉ bastian.gonzalezbustamante@politics.ox.ac.uk

Mesa-redonda: Estudos e configurações de elites na América Latina

45° Encontro Anual da ANPOCS

19 a 27 de outubro, 2021

1. Introducción
2. Confiabilidad/calidad de la información
3. Comparaciones y predicciones sesgadas
4. Regresiones de Cox, propensión y emparejamiento



Introducción

Cuatro problemas comunes identificados por [González-Bustamante \(2020\)](#) en [Olivares et al. \(2020\)](#):

- Estrategia de identificación.
- Confiabilidad/calidad de la información.
- Análisis de vínculos con redes sociales.
- Riesgo de comparaciones y predicciones sesgadas.

En esta presentación vamos a profundizar en la **confiabilidad/calidad de la información** y **predicciones sesgadas** asociadas con el enfoque cuantitativo.

Confiabilidad/calidad de la información

Existen problemas para obtener información confiable. Esto se debe principalmente a dos fenómenos: problemas de muestreo y deseabilidad social.

El **muestreo** constituye un desafío relevante al momento de diseñar un proyecto de investigación cuya unidad de análisis es la élite. Es complejo diseñar una muestra probabilística. Lo usual es intentar censar o usar muestras no probabilísticas reportando la tasa de respuesta. Sin embargo, bajas tasas de respuesta pueden ser un problema.

¿Lo aceptable? Sobre un 60 %. La validación de la información puede ser muy compleja debido a las estrategias de identificación.



UNIVERSIDAD
DE CHILE



UNIVERSIDAD
CATÓLICA
DE CHILE



udp



UNIVERSIDAD ADOLFo IBÁÑEZ



SERVEL

Encuesta a candidatos a Convencionales Constitucionales, Alcaldes y Concejales

RESULTADOS FINALES

Centro de Estudios
de Conflicto y
Cohesión Social



COES

Ficha técnica (I)

Estudio cuantitativo, elaborado mediante encuestas online utilizando la plataforma Qualtrics. Esta encuesta fue enviada a cada candidata o candidato para los distintos cargos, incluyendo al inicio del cuestionario un consentimiento informado.

La encuesta estuvo abierta a las distintas candidaturas durante 3 meses (26 de marzo - 26 de mayo 2021) y cuenta con un total de 3.035 casos (sobre un total de 16.553 candidatos y candidatas, exceptuando los candidatos a Gobernadores Regionales)

18,3%

5

Por otro lado, la **deseabilidad social** tiende a aumentar cuando se recaba información delicada sobre percepciones, lo que puede generar distorsiones en estudios de congruencia ideológica o identidad partidaria.

Los individuos suelen ocultar información o sobredimensionar sobre sus propias trayectorias.

Posibles soluciones:

- **Codificación manual.** Verificación minuciosa e intentar triangular distintas fuentes de información.
- **Codificación automática.** Algoritmos OCR y machine learning ([Courtney et al., 2020](#); [Radford, 2019](#), para un ejemplo véase [González-Bustamante et al., 2021](#)).

Comparaciones y predicciones sesgadas

El estudio de las élites no se encuentra exento de la influencia de la econometría en las ciencias sociales durante las últimas décadas. Sin embargo, existen altos riesgos de generar comparaciones y predicciones sesgadas.

Por ejemplo, para un caso i se necesita contar con una variable dependiente Y identificable que puede reflejar la trayectoria del caso con algún índice compuesto o indicadores más simples. En este contexto, es necesario contar con **varianza** en Y , ya sea una variable continua o dicotómica.

En otras palabras, si es binaria se necesitan casos positivos $Y = 1$ y negativos $Y = 0$. Esto es indispensable para ajustar modelos básicos como OLS o logits.

En algunas ocasiones es difícil contar con $Y = 0$. Por ejemplo, si se estudian nombramientos y no trayectorias o permanencia. No es posible contar con un conjunto de aquellos **individuos que fueron considerados para la posición**, pero no fueron nombrados (contrafactuales).

Algunas excepciones son posiciones de alta dirección pública en sistemas de servicio civil con concursos. También, evidentemente, los cargos de votación popular ya que hay registro de aquellos que perdieron la elección.

Una alternativa es con una muestra compuesta por distintas posiciones generar comparaciones entre grupos para detectar factores de nombramiento y después calcular probabilidades, sin embargo, es poco preciso.

Comparaciones y predicciones sesgadas

A lo anterior se suma, el **problema para desarrollar una estrategia de inferencia causal robusta**. Se puede considerar que el resultado de la inferencia causal es la diferencia en la variable dependiente Y para un caso i cuando es positiva y negativa: $Y_{1[i]} - Y_{0[i]}$ (Angrist y Pischke, 2015; Imai, 2017).

Entonces, el efecto de una variable independiente X , o de un tratamiento D (métodos experimentales), sería aquella diferencia. *E.g.* cuánto afecta X a Y en un monto κ (supuesto de efecto constante). Entonces es posible asumir que $Y_{1[i]} = Y_{0[i]} + \kappa$.

El problema fundamental de la inferencia causal es que **no es posible observar los resultados contrafactuales**. *E.g.* si X_i o D_i generan un cambio en Y_i , para estimar con precisión el efecto es necesario considerar los dos posibles potenciales resultados de Y_i , en este caso, $Y_{1[i]}$ e $Y_{0[i]}$. Es decir, **no solo se necesita varianza, son necesarios los contrafactuales** (e.g. ministros que fueron considerados y no nombrados).

Comparaciones y predicciones sesgadas

Es relevante considerar que Y no solo está determinada por X o D . Si Y es la carrera política de i , es posible identificar otros determinantes. *E.g.* el capital familiar permite a i tener ventajas debido a los procesos de cognición y socialización a los que estuvo expuesto.

¿Cuál factor es crítico? En este contexto, resulta complejo que en **datos observacionales** $D_i = 1$ y $D_i = 0$ se distribuyan de forma similar en los casos con diferentes resultados en la variable dependiente $Y_{1[i]}$ e $Y_{0[i]}$, lo que **impide la selección aleatoria**, bloqueando el camino metodológico más directo para eliminar el sesgo.

- *E.g.* Capital familiar: La mayoría de los casos exitosos lo tienen, por tanto, es altamente probable que con regresiones observacionales sea una variable significativa ya que no he eliminado el sesgo de selección.
- Lo mismo puede suceder con las credenciales académicas.

Probablemente un gran número de casos tenga capital familiar y sean exitosos, mientras que entre los casos no exitosos muy pocos tengan parentela política, por lo que se genera un **desbalance**. Resulta relevante eliminar el sesgo de selección que surge de diferencias no observadas:

$$Y_i = \alpha + \beta D_i + \gamma A_i + \epsilon_i \quad (1)$$

Dónde α es el equivalente a β_0 , mientras β y γ equivalen a β_1 y β_2 en la notación clásica de mínimos cuadrados. Por otra parte, D y A son equivalente a X_1 y X_2 , por tanto, su diferenciación es puramente conceptual y no formal. Por último, ϵ_i representa los residuos, lo que no se aborda para no complejizar el ejemplo.

Comparaciones y predicciones sesgadas

Cuando D_i es **asignado aleatoriamente**, la expectativa condicional de Y_i con $D_i = 1$ es equivalente a cuando $D_i = 0$, lo que puede ser complejo de obtener empíricamente con datos observacionales como se mencionó anteriormente:

$$\mathbb{E} [Y_i \mid D_i = 1] = \mathbb{E} [Y_i \mid D_i = 0] \quad (2)$$

Esta equivalencia permite una **medición adecuada del efecto causal** ya que se eliminan las diferencias sistemáticas entre grupos u observaciones como se aprecia a continuación:

$$\begin{aligned} & \mathbb{E} [Y_i \mid D_i = 1] - \mathbb{E} [Y_i \mid D_i = 0] \\ &= \mathbb{E} [Y_{0i} + \kappa \mid D_i = 1] - \mathbb{E} [Y_{0i} \mid D_i = 0] \\ &= \kappa + \mathbb{E} [Y_{0i} \mid D_i = 1] - \mathbb{E} [Y_{0i} \mid D_i = 0] \\ &= \kappa \end{aligned} \quad (3)$$

Comparaciones y predicciones sesgadas

PART I

AN EQUATION SATISFIED BY THE PROBABILITIES THAT A PLANE WILL BE DOWNED BY i HITS¹

INTRODUCTION

Denote by P_i ($i = 1, 2, \dots$, ad inf.) the probability that a plane will be downed by i hits. Denote by p_i the conditional probability that a plane will be downed by the i -th hit knowing that the first $i - 1$ hits did not down the plane. Let $Q_i = 1 - P_i$ and $q_i = 1 - p_i$ ($i = 1, 2, \dots$, ad inf.). It is clear that

$$Q_i = q_1 q_2 \dots q_i \quad (1)$$

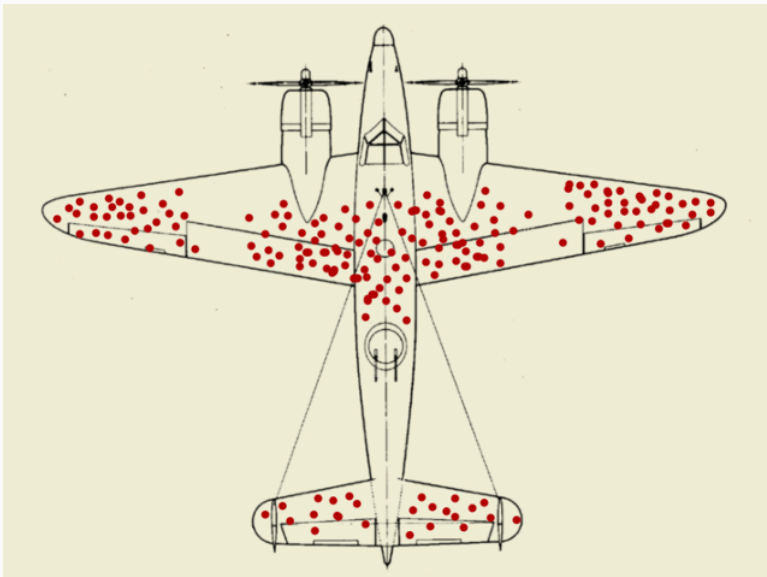
and

$$P_i = 1 - q_1 q_2 \dots q_i \quad (2)$$

Suppose that p_i and P_i ($i = 1, 2, \dots$) are unknown and our information consists only of the following data concerning planes participating in combat:

- The total number N of planes participating in combat.
- For any integer i ($i = 0, 1, 2, \dots$) the number A_i of planes that received exactly i hits but have not been downed, i.e., have returned from combat.

Comparaciones y predicciones sesgadas



Comparaciones y predicciones sesgadas

En consecuencia, si los datos no fueron asignados aleatoriamente, aunque la muestra sea probabilística, no se elimina el sesgo y, por tanto, las predicciones realizadas con modelos son imprecisas. Para solucionar es posible seguir alguno de estos caminos:

(i) Sacrificar el análisis probabilístico por uno determinista (QCA)

- Puede ser útil descriptivamente, particularmente para ver trayectorias y poscarreras. Aunque puede ser considerado poco ortodoxo.

(ii) Regresiones observacionales con estrategia de inferencia causal

- Técnicas de emparejamiento y control por indicadores de propensión (*propensity score*) o simulaciones de *pseudo*-observaciones para modelos OLS/logits/supervivencia.
- Adaptaciones más complejas y simulaciones para modelos de conteo y con distribuciones de Poisson.

Regresiones de Cox, propensión y emparejamiento

Regresiones de Cox con exposición variable en el tiempo

Para la estimación econométrica con datos observacionales, en [González-Bustamante \(2021\)](#) utilizo una extensión de los modelos de Cox con **riesgos no proporcionales** por la estructura del conjunto de datos. La ecuación incorpora el efecto de los intervalos $Z(t)$ de las múltiples observaciones i -th, considerándolas como *clusters*. Primero, estimamos un modelo con la variable de exposición trimestral a la **baja aprobación presidencial D_i (bajo 40 puntos)**. Controlamos efectos fijos por gobierno y país e incluimos un vector de controles C_k con k -th potenciales *confounders* (luego se utilizan para estimar *propensity scores*).

$$\lambda(t_i) = \lambda_0(t_i) \exp [\beta_t Z_i(t) + \beta D_i + \zeta_{gov_i} + \eta_{country_i} + \sum_{k=1}^K \vartheta_k C_{k[i]} + \varepsilon_i] \quad (4)$$

Regresiones de Cox con exposición variable en el tiempo

Luego, incorporamos el **efecto de interacción** entre la baja aprobación presidencial D_i y **covariables moderadoras** $X_{j[i]}$, denotando las covariables de tipo j -th que se incorporan en **modelos separados**. Nuestra estrategia consiste en probar tres efectos binarios de moderación relacionados con los perfiles y atributos de los ministros: (i) ministros no partidistas; (ii) economistas; y (iii) líderes de partidos.

$$\lambda(t_i) = \lambda_0(t_i) \exp [\beta_t Z_i(t) + \beta D_i + \gamma_j X_{j[i]} + \delta_j D_i \times X_{j[i]} + \zeta gov_i + \eta country_i + \sum_{k=1}^K \vartheta_k C_{k[i]} + \varepsilon_i] \quad (5)$$

El vector C_k ($K = 9$) considera las siguientes variables: **NEP en el legislativo**, una variable binaria que refleja cuando se permite la **reelección presidencial** consecutiva, **crecimiento económico**, **inflación** y cinco variables binarias que miden los patrones cuadráticos del **modelo cíclico** de la aprobación presidencial (*i.e.*, luna de miel inicial, un deterioro gradual y una leve recuperación al final mandato, véase [Carlin et al., 2018](#); [Stimson, 1976](#)).

Estas nueve covariables se consideran **potenciales confounders para la estimación de propensity score**, ya que podrían afectar tanto a la aprobación presidencial como a la decisión de despedir a los ministros.

- Para más información sobre los mecanismos causales de estas variables, véase el paper completo.

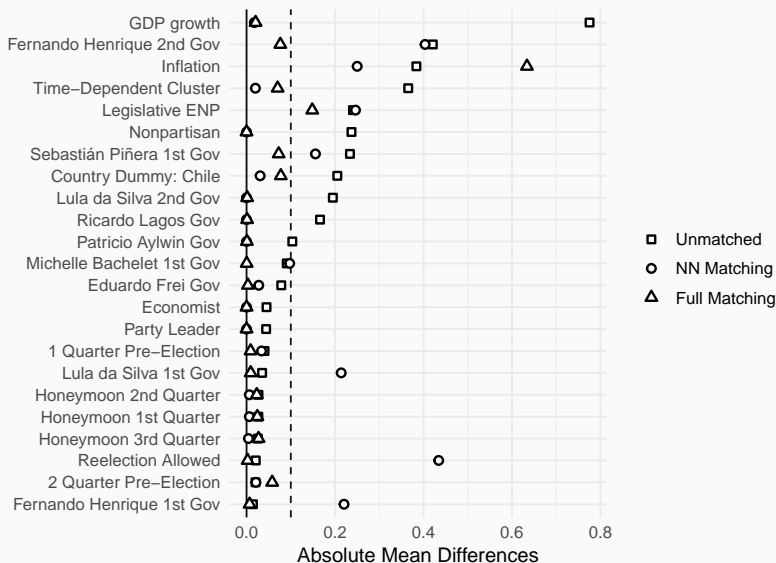
Análisis de propensión y emparejamiento

Para controlar el problema de la selección no aleatoria y el sesgo, empleamos PS y emparejamiento con una **regresión logística de períodos de baja aprobación presidencial** D_i con las covariables moderadoras $X_{j[i]}$, efectos fijos de gobiernos y de países, y los potenciales *confounders* C_k .

$$D_i = \alpha + \gamma_1 X_{1[i]} + \gamma_2 X_{2[i]} + \gamma_3 X_{3[i]} + \zeta_{gov_i} + \eta_{country_i} + \sum_{k=1}^K \vartheta_k C_{k[i]} + \varepsilon_i \quad (6)$$

Luego, clasificamos las observaciones *i-th* en diferentes grupos balanceados con PS y el algoritmo **nearest neighbour** sin reemplazo para emparejar las observaciones con y sin exposición a una aprobación baja D_i . También utilizamos **full matching** para obtener muestras con menor PS promedio entre observaciones pareadas, evitando descartar casos (Hansen, 2004; Olmos y Govindasamy, 2015).

Diferencias medias antes y después del emparejamiento



Usamos full matching e integramos NEP, que no se equilibra, como control en los modelos finales (Austin et al., 2007; Olmos y Govindasamy, 2015). Además, incorporamos las **ponderaciones obtenidas con el proceso de emparejamiento** y ajustamos los grupos *s-th* del emparejamiento para probar el efecto de la exposición a la baja aprobación presidencial en las salidas individuales de ministros.

$$\lambda(t_i) = \lambda_0(t_i) \exp [\beta_t w_i Z_i(t) + \beta w_i D_i + \mu w_i ENP_i + \xi w_i inflation_i + \epsilon_{i[s]}] \quad (7)$$

Posteriormente, incluimos términos de interacción entre la baja aprobación D_i con nuestros factores de moderación $X_{j[i]}$ en **modelos separados** para realizar el análisis de moderación.

$$\lambda(t_i) = \lambda_0(t_i) \exp [\beta_t w_i Z_{i(t)} + \beta w_i D_i + \gamma_j w_i X_{j[i]} + \delta_j w_i D_i \times X_{j[i]} + \mu w_i ENP_i + \xi w_i inflation_i + \epsilon_{i[s]}] \quad (8)$$

Efecto de la baja aprobación en el despido de ministros

	Model I	Model II
Low Approval (< 40%)	0.705*** (0.231)	0.857*** (0.194)
Matching Sub. Clustering	Before No	Full Yes
Moderation Covariates	Yes	PS
Legislative ENP	Yes	PS/Yes
Re-Election Allowed	Yes	PS
GDP	Yes	PS
Inflation	Yes	PS/Yes
Quadratic Approval Pattern	Yes	PS
Government FE	Yes	PS
Country FE	Yes	PS
Obs. Clustering	Yes	PS
Log-Rank	113.624***	22.129***
AIC	2696.23	1551.612
C-Index	0.672	0.58
VIF	1.084	3.441
Events	256	256
N	4245	4245
Log Likelihood	-1328.115	-772.806

Exponenciando los coeficientes β se obtiene el *hazard ratio* y se observa que el emparejamiento permitió una corrección del sesgo del 16,5 %: $e^{\beta} = 2,357$ ($p = 0,001$ y CI 95 %: 1.398 a 3.973). Esto implica un mayor riesgo de salida del gabinete de 135,7 % durante períodos de baja aprobación.

* $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$

Análisis de moderación de la exposición a la baja aprobación

	Time-Varying Cox Regressions			Survival Outcome Models		
	Model I	Model II	Model III	Model IV	Model V	Model VI
Low Approval ($< 40\%$)	0.877*** (0.240)	0.736*** (0.238)	0.619** (0.258)	1.389*** (0.232)	0.939*** (0.211)	0.674* (0.233)
Nonpartisan	-0.240 (0.215)			0.476 (0.261)		
Economist		-0.049 (0.207)			0.208 (0.301)	
Party Leader			0.241 (0.168)			-0.009 (0.262)
Low Approval x Nonpartisan	-0.771** (0.328)			-1.516** (0.362)		
Low Approval x Economist		-0.192 (0.336)			-0.392 (0.400)	
Low Approval x Party Leader			0.186 (0.272)			0.504 (0.339)
Matching Sub. Clustering	Before No	Before No	Before No	Full Yes	Full Yes	Full Yes
Legislative ENP	Yes	Yes	Yes	PS/Yes	PS/Yes	PS/Yes
Re-Election Allowed	Yes	Yes	Yes	PS	PS	PS
GDP	Yes	Yes	Yes	PS	PS	PS
Inflation	Yes	Yes	Yes	PS/Yes	PS/Yes	PS/Yes
Quadratic Approval Pattern	Yes	Yes	Yes	PS	PS	PS
Government FE	Yes	Yes	Yes	PS	PS	PS
Country FE	Yes	Yes	Yes	PS	PS	PS
Obs. Clustering	Yes	Yes	Yes	PS	PS	PS
Log-Rank	125.147***	98.495***	105.73***	49.198***	23.145***	29.188***
AIC	2690.52	2708.627	2703.748	1532.641	1554.651	1550.508
C-Index	0.683	0.663	0.668	0.657	0.607	0.623
VIF	1.081	1.082	1.078	3.452	3.467	3.444
Events	256	256	256	256	256	256
N	4245	4245	4245	4245	4245	4245
Log Likelihood	-1326.260	-1335.313	-1332.874	-761.321	-772.325	-770.254

* $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$

Los resultados principales nos permiten comprobar que la probabilidad de destitución ministerial aumenta durante los períodos de baja aprobación: el riesgo se duplica, aumentando en un 135,7 %.

Además, **los ministros no partidistas están más protegidos durante estos períodos**: sólo uno de cada cinco ministros no partidistas es destituido del gabinete en una crisis de aprobación en comparación con los ministros partidistas.

Con el emparejamiento se controló los desequilibrios y sesgos, logrando una **corrección del 16,5 % y del 52,5 %** para el efecto principal y el de moderación, respectivamente.

Referencias I

- Angrist, J. D. y Pischke, J.-S. (2015). *Mastering Metrics. The Path from Cause to Effect*. Princeton University Press, Princeton.
- Austin, P. C., Grootendorst, P., y Anderson, G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Statistics in Medicine*, 26(4):734–753.
- Carlin, R. E., Hartlyn, J., Hellwig, T., Love, G. J., Martínez-Gallardo, C., y Singer, M. M. (2018). Public support for Latin American presidents: The cyclical model in comparative perspective. *Research & Politics*, 5(3):1–8.
- Courtney, M., Breen, M., McMenamin, I., y McNulty, G. (2020). Automatic translation, context, and supervised learning in comparative politics. *Journal of Information Technology & Politics*, 17(3):208–217.
- González-Bustamante, B. (2020). El estudio de las élites políticas gubernamentales en América Latina: Panorama, agendas de investigación y desafíos metodológicos. SocArXiv DOI: 10.31235/osf.io/syqu4.
- González-Bustamante, B. (2021). Ministerial Stability During Presidential Approval Crises: The Moderating Effect of Ministers' Attributes on Dismissals in Brazil and Chile. En *XXVI World Congress of Political Science*, Lisboa. International Political Science Association.

Referencias II

- González-Bustamante, B., Astete, M., y Orvenes, B. (2021). Senior Public Managers: A Novel Dataset on Members of the Chilean Civil Service. SocArXiv DOI: 10.31235/osf.io/vshcz.
- Hansen, B. B. (2004). Full Matching in an Observational Study of Coaching for the SAT. *Journal of the American Statistical Association*, 99(467):609–618.
- Imai, K. (2017). *Quantitative Social Science: An Introduction*. Princeton University Press, Princeton.
- Olivares, A., González-Bustamante, B., Toro, S., Arellano, J. C., Yanes-Rojas, A., Zurita-Tapia, J., Lopes, A. V., Robelo Guzmán, C., y Canavesi Sosa, J. B. (2020). Nuevos desafíos, enfoques y perspectivas para estudiar élites políticas. *Iberoamericana. América Latina - España - Portugal*, 20(74):229–259.
- Olmos, A. y Govindasamy, P. (2015). Propensity Scores: A Practical Introduction Using R. *Journal of MultiDisciplinary Evaluation*, 11(25):68–88.
- Radford, B. J. (2019). Automated dictionary generation for political event coding. *Political Science Research and Methods*, 9(1):157–171.
- Stimson, J. A. (1976). Public Support for American Presidents: A Cyclical Model. *Public Opinion Quarterly*, 40(1):1–21.

Bastián González-Bustamante

DPhil (PhD) Researcher

Department of Politics and International Relations
& St Hilda's College
University of Oxford

📍 St Hilda's College, Cowley Place, Oxford OX4 1DY

✉ bastian.gonzalezbustamante@politics.ox.ac.uk

🏠 <https://bgonzalezbustamante.com>

Profesor Instructor

Departamento de Gestión y Políticas Públicas
Facultad de Administración y Economía
Universidad de Santiago de Chile

📍 Av. Lib. B. O'Higgins 3363, Estación Central, Santiago

✉ bastian.gonzalez.b@usach.cl



Presentación compilada con \LaTeX y algunos 📄

🔄 Descargar la versión más reciente desde [GitHub](#)

Muchas gracias por su atención

